

DRAFT *VERSION 1.0*

NEWDISS



A 6-TO 10-YEAR APPROACH TO DATA SYSTEMS AND SERVICES FOR NASA 'S EARTH SCIENCE ENTERPRISE

COVER ILLUSTRATION:

In a manner corresponding to the way that neural complexity is built upon simple physiological types organized into greater and greater hierarchies of complexity, NewDISS is founded on a set of simple functions that interact according to a set of established criteria, organized by individual and institutional partners into hierarchies of functional groupings optimized to perform specific tasks, the whole of which functions together to support the sentient goals of the Earth Science Enterprise.

EXECUTIVE SUMMARY

What is NewDISS?

NewDISS is the NASA infrastructure that links existing data resources, while facilitating the evolution of data management into information and knowledge management. NewDISS integrates Distributed Active Archive Systems (DAACs), the EOSDIS Core System (ECS), and the Earth Science Information Partnerships (ESIPs) by adding data handling systems that simplify discovery, access, and manipulation of NASA data sets. NewDISS facilitates development of advanced information and knowledge management systems by providing standard interfaces for describing information (attributes about data) and knowledge (relationships between attributes). NewDISS manages evolution by interconnecting heterogeneous data storage and information repository resources through use of published standards. NewDISS will provide the technology evolution management needed to ensure continued ability to use NASA data to explore Earth Systems Science research. This includes the ability to exploit data collections to meet new scientific challenges.

What's New in NewDISS?

NewDISS refers to the distributed Earth science data systems and services, which, over the next 6–10 years, will evolve after the Earth Observing System Data Information System (EOSDIS). Its prime goal will be to support NASA's Earth Science Enterprise (ESE), which contributes, in turn, to the U.S. Global Change Research Program (USGCRP). As such, NewDISS is driven principally by objectives of scientific research.

NewDISS will consist of a heterogeneous mix of interdependent components derived from the contributions of numerous individuals and institutions. These widely varying participants will be responsible for data management functions including data acquisition and synthesis, access to data and services, and data stewardship. Because the NASA ESE already has made considerable investment in existing data system activities (e.g., Distributed Active Archive Systems (DAAC's), the EOSDIS Core System (ECS), the Earth Science Information Partnerships (ESIP's)), and in product generation (Pathfinder Data Sets, mission data processing systems), the near-term NewDISS will necessarily leverage off of these existing components. However, the future NewDISS components could be quite different, as data systems and services evolve to meet science-driven demands and to take advantage of technological innovation.

A key goal of NewDISS is to harmonize and aggregate the various, disparate, and numerous data systems and services of NASA's ESE. In doing so, NewDISS will be built on a number of existing and evolving systems and services. As part of the NewDISS development, NASA must address the areas where improvements are required, and any principle of building on success means identifying success. Thus, success criteria and metrics will be an increasingly significant component of NewDISS management. From this perspective, NewDISS will adopt mechanisms for rapidly incorporating feedback from the science community on the design, deployment, and performance of the NewDISS services on an on-going basis.

NewDISS will need to be heterogeneous, with different, and perhaps more expandable or replaceable, nodes than we have today. **The concept of a reconfigurable network of services will be central to the NewDISS.** Consequently, NewDISS must be very simple, flexible and adaptable.

NewDISS Recommendations

Throughout the text of this report there are a number of recommendations made which are intended to help guide NASA in the development of the new data systems, services and processes that will make up NewDISS. We recommend that NASA must:

- **Support a spectrum of heterogeneous participants and approaches to NewDISS.** Such diversity should especially be considered in the investigators, organizations and institutions that make up NewDISS.
- **Support a spectrum of heterogeneous technological approaches to NewDISS.** NewDISS management must concentrate on integrating suitable existing data service capabilities, while also identifying and providing a means for delivering capabilities that do not yet exist.
- **Clearly define the components of NewDISS, and ensure suitable management of the interfaces between them.** This includes the definition of a set of "core" standards and practices, along with the means for selecting and maintaining them.
- **Employ a NewDISS infrastructure that includes active liaison with service providers both within NASA and within the private sector for procurement of common operations activities.**
- **Employ competition and peer review in the process used for choosing NewDISS components.**
- **Empower science investigators with an appropriate degree of responsibility and authority for NewDISS data system development, processing, archiving and distribution.**
- **Use lessons learned from the current, experimental ESE federation as a step towards the NewDISS, and proceed with the Federation Experiment with this evolution in mind.**
- **Charter, without delay, a transition team with the objective of developing a transition plan, based on the findings and recommendations of this document that would lead to the initiation of a NewDISS starting in 2001.** The transition plan should further elaborate the functions and cost estimates of NewDISS components, including the various functional data centers, and the infrastructure components necessary for a complete NewDISS. The plan should also establish criteria for choosing and evaluating the investigators, institutes and organizations that participate in NewDISS, and determine precisely which functions of

NewDISS must be centralized and which must not. Upon agreement of a formulation plan for NewDISS, NASA should quickly move to implementation.

TABLE OF CONTENTS

PREFACE.....	8
NEWDISS TEAM MEMBERS.....	9
A. INTRODUCTION AND OVERVIEW.....	10
A.1 The Rationale for New Data and Information System and Services	10
A.2 Antecedents of NewDISS	10
A.3 The Scope of NewDISS	11
A.4 Lessons Learned	13
A.5 The Science Research Strategy and NewDISS.....	14
A.6 NewDISS and Applications.....	15
A.7 A New Data and Information Systems and Services Approach	16
B. PRINCIPLES FOR NEW DATA INFORMATION SYSTEM AND SERVICES	18
C. The COMPONENTS OF NEWDISS	21
C.1 Current EOSDIS Infrastructure and NASA ESE Data and Information-Related Assets.	21
C.2 Proposed Near-Term NewDISS Components	23
C.3 An Implementation Scenario.....	26
D. NewDISS INTERFACES, STANDARDS AND PRACTICES.....	29
D.1 Core and Community Standards.....	30
D.2 Core Standards and Practices.....	31
D.3 NewDISS Core Interface Selection and Maintenance Process.....	32

E. MANAGEMENT AND LEADERSHIP.....	35
E.1 Managing Diversity and Integration	36
E.2 Structures for Management and Leadership	36
E.3 NewDISS Partnerships	38
E.4 Cost Considerations	40
E.5 Metrics for Success	42
E.6 Transition Planning	42
F. REFERENCES.....	43
APPENDIX A... ..	44
APPENDIX B.....	46
APPENDIX C–LIST OF ACRONYMS.....	50

PREFACE

In July 1998, Dr. Ghassem Asrar, NASA's Associate Administrator for Earth Science, constituted a "New Data and Information Systems and Services" (NewDISS) Strategy Team. This team was given the charter to define the future direction, framework, and strategy of NASA's Earth Science Enterprise (ESE) data and information processing, near-term archiving, and distribution.

In his charter to the NewDISS Strategy Team, Dr. Asrar noted that the largest single contract for ESE data management—the Earth Observing System Data Information System (EOSDIS) Core System—has been unable to keep pace with technology advances and programmatic changes. Consequently, Dr. Asrar asked the NewDISS Strategy Team to advise and recommend new processes, procedures, and methods for securing and providing data and information services. The charge to the team was to provide a long-term plan, recommending how to proceed beyond 2000 and throughout the next 6–10 years. While acknowledging the present need in information technology to create distributed, flexible, and responsive systems, and ESE's need to have smaller, more manageable pieces, Dr. Asrar recognized the necessity for a framework that integrates the ESE data and information activities.

This long-term plan is intended to answer this question: Based on "lessons learned," both in the immediate past of the ESE, and from all quarters inside and outside the Government, what is the recommended viable and evolvable way of building a set of data and information systems and services to meet the ESE program needs?

NEWDISS TEAM MEMBERS

Members

MARTHA E. MAIDEN, NASA Headquarters, *Chair*
VANESSA GRIFFIN, NASA Goddard Space Flight Center
MATHEW SCHWALLER, NASA Goddard Space Flight Center
CANDACE CARLISLE, NASA Goddard Space Flight Center
RONALD L. S. WEAVER, University of Colorado
ROY JENNE, National Center for Atmospheric Research
KAREN WHITE, NASA Headquarters
SARA J. GRAVES, University of Alabama, Huntsville
DAVID L. SKOLE, Michigan State University
ANNGIENETTA R. JOHNSON, NASA Headquarters
GUENTER R. RIEGLER, NASA Headquarters
THOMAS A. LASINSKI, Lawrence Livermore National Laboratory
JOHN R. G. TOWNSHEND, University of Maryland, College Park

Consultants

MARK R. ABBOTT, Oregon State University
GEORGE DAVID EMMITT, Simpson Weather Associates Incorporated
JAMES FREW, University of California, Santa Barbara
DAVID M. GLOVER, Woods Hole Oceanographic Institution
ANTHONY C. JANETOS, World Resources Institute
THOMAS KARL, NOAA National Climatic Data Center
PAMELA A. MATSON, University of California
DOROTHY PERKINS, NASA Goddard Space Flight Center
MOSHE PNIEL, Jet Propulsion Laboratory
CARL REBER, NASA Goddard Space Flight Center
RICHARD B. ROOD, NASA Goddard Space Flight Center
CHRISTOPHER SCOLESE, NASA Goddard Space Flight Center

The NewDISS Team authored Draft Version 0.1 of this report, published in July 1999, and Draft Version 0.2, published in October 2000. This Version 1.0 is the result of additions and edits for clarification prepared by NASA's Office of Earth Science in consultation with its ESSAAC {Earth System Science and Applications Advisory Committee} Subcommittee for Information Systems and Services (ESISS).

A. INTRODUCTION AND OVERVIEW

A.1 The Rationale for New Data and Information System and Services (NewDISS)

Satellites launched over the last four decades of the twentieth century have provided all-encompassing views of Earth within a single frame, thereby granting humanity its first look at the Earth in its totality. From the perspective of space, political boundaries are invisible; people of all nations and cultures can appreciate their shared, common environment. The extremely thin atmosphere that separates that environment and all life within it from the dark, cold emptiness beyond provides an implicit warning of how fragile our existence is.

NASA's Earth Observing System (EOS) builds on the foundation of global observation in an effort to build global understanding. It provides the capability to monitor continuously and comprehensively the health of the entire biosphere. Coupled with this scientific measurement capability is the explosive increase in technology for disseminating information, specifically via the internet and the world wide web. The web is revolutionary not just in its ability to deliver information anywhere; its grandest potential is that, by permitting information exchange throughout a global community, it can help synthesize global experience, understanding and knowledge. A third technology is interwoven with these two: geographic information systems (GIS) create a visual, spatial language, one that is inherently international. Maps and related spatial visualizations empower people to analyze their world in new ways.

The twenty-first century, therefore, begins with the promise of a world whose health is quantifiable, and a world that is globally networked and geographically transparent. It is within this context that NASA's New Data and Information Systems and Services (NewDISS) is proposed. It is the goal of NewDISS to build a knowledge base enabling as many people as possible to:

- Assess the health of the planetary environment;
- Predict changes that might occur to that environment;
- Judge whether those changes will be beneficial or detrimental; and
- Initiate actions that will mitigate detrimental changes and enhance beneficial ones.

A.2 Antecedents of NewDISS

The need for the development and implementation of distributed, open, extensible systems has long been recognized. A National Research Council (NRC) Committee on Data Management and Computation (CODMAC) report published in 1986 (NRC 1986) stressed the need for geographically distributed systems that should develop in an evolutionary fashion. In 1988 the same Committee recommended a distributed archive managed by each discipline. In 1991 the Committee on Geophysical and Environmental Data (CGED) of NRC argued that investment in data and information management should be visibly driven and accountable to the scientific objectives of the USGCRP (NRC 1991). There was also an emphasis on the need for an evolving data management

system. The report of the Committee on Geophysical and Environmental Data (CGED) commenting on a Federal Plan for Managing Global Change Data and Information, struggled with the issue of how the resultant Global Change Data and Information System will be managed and by whom (NRC 1992) prior to the maturing of ideas of a more federated approach to governance. In 1993 the same Committee again stressed the need for an evolving system that would be flexible and promote access for new users as well as enhancing services for established users (NRC 1993). The NRC Panel to Review EOSDIS Plans expressed concerns because the design was centralized and would likely be unable to keep up with the inevitable changes in technology and user needs over time (NRC 1994). In the Review of NASA's ESE Research Strategy NRC stressed the need for developing strategies for how NASA will deliver data and information to the broader scientific community (NRC 2000). The vision for NewDISS outlined in this document responds to many of these concerns for a more distributed system, while at the same time maintaining sufficient overall organizational structure to allow effective management and implementation.

A.3 The Scope of NewDISS

NewDISS specifically refers to the distributed Earth science data systems and services that, starting in 2000 and throughout the next 6–10 years, will evolve after the Earth Observing System Data Information System (EOSDIS). Its prime goal will be to support NASA's Earth Science Enterprise (ESE), which contributes, in turn, to the U.S. Global Change Research Program (USGCRP). As such, NewDISS is driven principally by objectives of scientific research. What this means in practice is that scientists in the pursuit of their goals should be able to gain access to data and information and the tools for analyzing and processing them, which facilitate to the greatest extent possible an improved understanding of the functioning of the Earth system.

Although the science goals of the Earth Science Enterprise will be the prime driver, NewDISS must be sufficiently flexible and capable of providing support to such other activities as NASA's Applications program and its educational and outreach programs. Thus, NewDISS must be capable of providing data and information not only from NASA's remote sensing missions but must also be capable of handling, integrating and distributing data from other key U.S. and overseas missions and from in situ observations.

This document develops the NewDISS concept in four major sections: starting with a set of principles, building a case for the components of NewDISS, defining the mechanisms by which these components are connected, and ending with a discussion of NewDISS management.

The principles of NewDISS are based on several simple and irrefutable observations. First, in the next decade NASA should organize its data systems to answer science questions and priorities. Second, NASA's Earth science data volume and user demand will increase over time. Third, technological change is continuous and inevitable. Fourth, competition is a key tool for selection of NewDISS components and infrastructure. Fifth, Principal Investigator data processing and data management will be a significant part of future ESE missions and

science. Sixth, long-term stewardship and archiving of ESE data must occur. And finally, data system implementation must be designed to evolve strategically.

Building on the principles outlined above, the elements of NewDISS are defined in Section C, along with a scenario for NewDISS implementation. In this conceptualization, NewDISS is described as a framework of distinct but cooperative components. These components handle the data management needs of ESE, and they also permit easy extensibility beyond the ESE per se.

Section D establishes the mechanisms by which the components of NewDISS are connected. In a fully mature NewDISS, with publicly published interface standards and practices, will allow inclusion of any who wish to participate by utilizing these standards and practices.

Section E of this document lays out the plans for the management, leadership and governance of NewDISS. In short, NASA will proceed with a distributed management paradigm, but the agency will provide NewDISS integrity through a leadership role. NASA will define priorities and practices for NewDISS with the goal of establishing a standard approach and rational set of criteria for evaluating the success (or lack of it) of investigators, institutes and organizations that participate in NewDISS. In doing so, any failures of NewDISS can be quickly identified and corrected, and the successes of NewDISS can be identified and encouraged.

As described in the report below, the proposed NewDISS structure is built on the concept of distributed, "focused" mission and science data centers which will leverage on-going data management activities at educational institutions, NASA and other government organizations, and in the private sector. It is also anticipated that the NewDISS will be highly distributed, with many of its functions being conducted within laboratories run by Principal Investigators.

Practical considerations drive initial participation and evolution of NewDISS. Therefore, it is already possible to identify some of the existing components and capabilities that will contribute to the early NewDISS. There will be parts of EOSDIS that will contribute, though without the centralization and constraints of this system. The existing DAAC's and their successors are expected to provide archiving capabilities. There are multiple Principal Investigator (PI) heritage instrument processing systems that can contribute to new mission systems. Additionally it is anticipated that PI's whose research leads to the creation of higher order products will also contribute to NewDISS, including, for example, the Pathfinder Data Set investigator systems and some of the Earth System Information Partnerships (ESIP's). The initial participation in NewDISS will evolve based on changing priorities and science questions.

An important premise underlying the operation of NewDISS is that its various parts should have considerable freedom in the ways in which they implement their functions and capabilities. Implementation will not be centrally developed, nor will the pieces developed be centrally managed. However, every part of NewDISS should be configured in such a way that data and information can be readily transferred to any other. This will be achieved

primarily through the adoption of common interface standards and practices. Another important premise is that users will readily be able to find the location of and gain access to the data sets that they need to achieve their scientific goals.

A.4 Lessons Learned

In the past, NASA has responded to the challenges of Earth science data management by focusing on a strategy of centralized data set production and distribution. The primary goals of such data systems were predictable schedules and standard data products. This strategy was motivated by an interest in minimizing the risk associated with developing new ground systems for Earth science spaceflight missions: the ground data system was viewed as an extension of the spacecraft. If the spacecraft could be engineered to schedules and specifications, why not the ground data system? This approach worked reasonably well, but suffered under the pressures of data management requirements for the EOS satellites. Recent evaluations of NASA's data management practices have yielded five key "lessons learned." These are based on observation of NASA's EOSDIS and other large-scale software development efforts, and are corroborated by NASA's Office of Space Science (see Appendix B).

- **Information technology outpaces the time required to build large, operational data systems and services.** Technology is now changing at such a rapid pace that it is impossible to predict technology solutions even 2 years into the future. And, in contrast to 10–15 years ago, government information systems no longer drive the business of hardware and software vendors; NASA is now just another customer trying to capture the attention of the vendors.

- **Data systems and services should leverage off emerging information technology, and not try to drive it.** As described above, NASA is no longer solely in a position to drive commercial hardware and software development. Thus, NewDISS must be open to the infusion of new technologies developed by industries that a few years ago were completely unassociated with digital information management, but are now leaders in the field, e.g., the banking, entertainment and retail industries.

- **A single data system should not attempt to be all things to all users.** The ESE research and applications community is extraordinarily diverse, with interests extending from the top of the atmosphere to the Earth's core. The standards and practices governing the acquisition, archiving, documentation, distribution, and analysis of Earth science are, de facto, those established by the disciplines specific scientific peer groups within this community. An ESE NewDISS must recognize and embrace this tapestry of disciplines and subcommunities; there is no one-size-fits-all solution to the myriad data management needs of the community as a whole.

- **A single, large design and development contract stifles creativity.** Given the complexity of the required systems and services, the volatility of the technology, and the potential for changes in scientific priorities, centralized development is too inflexible and increases the risk that large portions of the data system will be vulnerable to single-point failures. Such an

approach is also prone to “monopolistic” tendencies, and does not encourage the kind of diversity and variety found in a competitive marketplace.

- **Future information systems will be distributed and heterogeneous in nature.**

Structuring NewDISS around science questions and the needs of the science community suggests that the science community must be involved directly in its overall design and implementation. This implies a reliance on a highly distributed and heterogeneous network of producers and consumers. Furthermore, management tools and practices must encourage a flexible, distributed, and loosely coupled network of data producers and providers, even if this requires a fundamentally new management approach within the NASA culture.

In the future substantially higher volumes of data will be used resulting from the need for increasingly finer resolution analyses and models, the need to apply multiple data sets from many sources as well as the multiple needs of an ever-widening user base. Today’s Earth observation satellites can generate data at the rate of 10–20 Mbits per second continuously over a 6- year lifetime. Plus, EOS data systems need to archive and distribute a host of derived and ancillary data products, driving the total mission data volumes several times higher than the raw data rates. Also, the user community for these data is diverse, and is expected to become larger and ever more diverse over time. This community can be expected to demand ever more complex systems and services as it attempts to mine a diverse archive of data generated by current and planned Earth observation satellites and related programs.

The goal of the NewDISS is to assure that an adequate data service capability exists to meet the current suite of Earth science data management challenges. This does not mean that NewDISS should assume the responsibility for building a comprehensive data center or a data system to meet all the needs of all potential users. Rather, the NewDISS must provide a flexible framework to integrate data service capabilities, common interfaces both from within and outside of NASA. NewDISS must also identify needed capabilities that are not now available and must facilitate the development of those capabilities. In doing so, NewDISS will be continually changing, continually seeking the goal of optimizing performance and usability for the ever-changing aggregate of ESE data activities. The remainder of this document provides an approach to reaching this goal.

A.5 The Science Research Strategy and NewDISS

In the NASA Earth Science Enterprise Research Strategy for 2000-2010 (NASA 2000) the ESE has defined its research strategy around a hierarchy of scientific questions arranged within a framework of five steps or fundamental questions:

- How is the global Earth system changing?
- What are the primary forcings of the Earth system?
- How does the Earth system respond to natural and human-induced changes?
- What are the consequences of change in the Earth system for human civilization?
- How well can we predict future changes in the Earth system?

This represents a fundamental shift from the previous thematic approach. The previous decade's premier Earth Observing System program was highly focused on variability, and the posed questions shaped a strategy that was centered on the generation of 24 key sets of observations.

NewDISS will place NASA in a much better position to address the next decade's essentially inter-disciplinary questions. It will lead to multi-mission and science-based centers targeted respectively at the integration of data sets and specific research questions. Moreover, with its emphasis on interfaces between centers and other components of the DISS, the integration and use of multiple data sets aimed at answering science questions should be considerably enhanced.

The much more flexible approach adopted in NewDISS will allow it better to meet the scientific challenges of the next 10 years. In terms of data there will be the demand for larger volumes of data associated with higher resolution modeling and prediction as well as the need to draw upon multiple types of observations. The increasingly close interactions between observations, their analysis and reanalysis also leads to the need for systems much more closely integrated with the scientific community.

NewDISS will also have to meet the challenge of creating consistent systematic measurements for the measurement and detection of long-term change. Experience indicates that with the use of multiple observing systems considerable scientific involvement is needed to achieve high quality long-term data sets.

A.6 NewDISS and Applications

Science requirements served as the principal design driver behind the present-day EOSDIS. In the future, applications requirements are envisioned to play a significant role in the design and operation of NewDISS.

The applications goal of the ESE is to **expand and accelerate the realization of economic and societal benefits from Earth science, information, and technology**. This goal has served as the driver for the establishment of the Applications program within ESE. "Applications" are defined as either new information goods of practical value, or new uses for data, information or technology originally developed for scientific research. The applications program is managed by the Applications Division, which has the mission to turn scientific and technical capabilities into practical tools for public- and private-sector decision makers.

The Applications program is organized around four themes: resource management, disaster management, community growth and infrastructure, and environmental assessment. These four themes are the primary framework around which sponsored activities are organized. These themes also serve as a meaningful system by which to segment applications NewDISS users, many of whom may be pursuing applications development investigations independent of NASA sponsorship within either the public or private sector. It is also anticipated that some of these users that achieve a sustainable,

routine usage of geospatial information goods may interact with the system in an operational or semi-operational context.

Applications users will form a distinct class of NewDISS customers and participants. As customers, applications users may serve as consumers of "geospatial information goods", derived from ESE observations and models, distributed through NewDISS. Many applications users may be strictly end-user consumers of NewDISS data, information, or services. They may be part of either the private or public sector. As participants, applications users may also serve as producers of geospatial information goods that will be provided to and distributed by NewDISS. Applications users that are producers may be part of commercial enterprises that produce remote sensing data on their own from either spaceborne or airborne platforms or they may be part of the satellite remote sensing "value-added industry". Other applications users that are producers may be from the public sector at the national, state, local, or tribal level, or may be part of non-governmental organizations.

The NewDISS should be responsive to these distinct user groups, accommodating their needs and allowing sufficient extensibility so that new producers be allowed to "plug in" as desired.

At the highest level, NewDISS levels of service for applications users should help to break the traditional barriers to the proliferation of remote sensing data, information, services or technology. These barriers include cost, access, and appropriateness of the data to the problem at hand. Examples of features of NewDISS that may ease access barriers include easy-to-use user interfaces, online data staging, and ability to perform content-based querying. Examples of features of NewDISS that may make the data more appropriate to the problem at hand include interactive subsetting, interactive data reformatting, convenient metadata representations, and other on demand "tailoring" of geospatial data products.

Finally, as the system evolves, it must be considered that new applications requirements may appear that contend with or reach beyond requirements of the science user community. NewDISS must include processes by which these types of challenges can be answered.

A.7 A New Data and Information Systems and Services Approach

NewDISS refers to the distributed Earth science data systems and services that will evolve over the next 6–10 years. Its prime goal will be to support NASA's Earth Science Enterprise (ESE), which contributes, in turn, to the U.S. Global Change Research Program (USGCRP). As such, NewDISS is driven principally by objectives of scientific research.

NewDISS will consist of a heterogeneous mix of interdependent components derived from the contributions of numerous individuals and institutions. These widely varying participants will be responsible for data management functions including data acquisition and synthesis, access to data and services, and data stewardship. Because the NASA ESE already has made considerable investment in existing data system activities (e.g., Distributed Active Archive

Systems (DAACs), the EOSDIS Core System (ECS), the Earth Science Information Partnerships (ESIPs)), and in product generation (Pathfinder Data Sets, mission data processing systems), the near-term NewDISS will necessarily leverage off of these existing components. However, the future NewDISS components could be quite different, as data systems and services evolve to meet science-driven demands and to take advantage of technological innovation.

A key goal of NewDISS is to harmonize and aggregate the various, disparate, and numerous data systems and services of NASA's ESE. In doing so, NewDISS will be built on a number of existing and evolving systems and services. As part of the NewDISS development, NASA must address the areas where improvements are required, and any principle of building on success means identifying success. Thus, success criteria and metrics will be an increasingly significant component of NewDISS management. From this perspective, NewDISS will adopt mechanisms for rapidly incorporating feedback from the science community on the design, deployment, and performance of the NewDISS services on an on-going basis. NewDISS will need to be heterogeneous, with different, and perhaps more expandable or replaceable, nodes than we have today. **The concept of a reconfigurable network of services will be central to the NewDISS.** Consequently, NewDISS must be very simple, flexible and adaptable.

B. PRINCIPLES FOR NEW DATA INFORMATION SYSTEMS AND SERVICES

This section presents the overarching principles upon which NewDISS will be developed and provides a basis for the discussion in later sections. Consistent with the NewDISS concept of flexible, distributed system elements, which have considerable implementation freedom, this report focuses on the concepts for a NewDISS, rather than on explicit implementation specifications. It is therefore crucial that NewDISS adhere to some common principles underpinning any specific suite of systems or services.

The principles for NewDISS start from a premise (stated in section A.1) that systems and services must be informed by and supportive of key science concerns and questions. From here we recognize that individual scientists as well as disciplinary communities of scientists become the key consumers and producers of data products and derived information, and therefore must be key partners. Other key principles relate to the issue of immediate and long-term services for a highly distributed and heterogeneous user base in the face of rapid technological change. These key principles are summarized as follows:

- **Science questions and priorities must determine the design and function of systems and services.**
- **Future requirements will be driven by a high data volume, a broader user base and increasing demand for a variety of data and data products.**
- **Technological change will occur rapidly: systems and services must be able to take advantage of these changes.**
- **Competition is a key NASA tool for selection of NewDISS components and infrastructure.**
- **PI-processing and PI data management will be a significant part of future missions and science.**
- **Long-term stewardship and archiving must occur.**
- **NewDISS evolutionary design must move beyond data and information and towards knowledge-based systems.**

Science questions and priorities must determine the design and function of systems and services. In the past, ground data systems were designed to meet requirements for processing raw satellite data through to finished science data products. With this approach, considerable resources and management emphasis were placed on the few critical data facilities needed to acquire the raw satellite data and generate products. Typically, these systems were developed at NASA centers or other government institutions. As such, the data centers were developed and located outside the science community, while providing data support to the science community. This approach, based on a relatively static set of requirements, was not designed particularly well for responsiveness to changes in direction motivated by the science community. This lack of agility becomes increasingly important since a recent review from the National Research Council recommended restructuring Earth-observation activities, with a new orientation toward unanswered scientific questions (NRC 1999). The implication of this recommendation is that as resources get redirected toward new scientific priorities, data systems and services must be capable of responding quickly, without radical and costly

redevelopment of hardware and software. Thus, as science questions change, the systems and services will need to be flexible and easily changed too, substituting or redirecting the aims of various data centers, replacing or eliminating unnecessary services and centers, and expanding or shrinking the entire system as the overall demand changes with the progress of the science and the programs. Clearly, NewDISS will need to permit successful centers to continue, grow or evolve, while also permitting less successful ones to be descope, eliminated, or replaced. This being said, the pace of change for facilities that acquire raw satellite data (and those that archive key historical records) is expected to be slower than that experienced in facilities that deal with more rapidly-changing data product suites.

Competition is a key NASA tool for selection of NewDISS components and infrastructure. Peer-review has long been the mainstay of the ESE community for achieving its technical and scientific objectives. Although recognized as not perfect, competition and peer review are widely and generally accepted as the best methodologies for maintaining excellence, containing costs, and providing an open environment for participation of the most capable. It is strongly recommended that NASA employ competition and peer review for all components of NewDISS. It is expected to be the case that some functions are more slowly varying than others. Therefore, competitions should be managed to take place along various appropriate timelines.

PI processing and PI data management will be a significant part of future ESE missions and research. The science community has managed data for a long time, but historically used different approaches across the various disciplinary communities, or through clusters of investigators centered on specific scientific problems. As NASA moves toward a series of missions that encompass both systematic and short-term observations, it will increasingly emphasize PI processing and PI data management as a tool for the rapid development and implementation needed to address critical science questions. In this environment, the time to implement science missions and deliver credible scientific information must be reduced. Therefore, it will be in the interest of the PI to deliver information rapidly to the broader community if the PI wishes to propose subsequent satellite missions. Rapid data delivery builds a strong community of support for the mission. One of the implicit assumptions in EOSDIS was that PI's needed to be coerced to provide data to the broader community. This is no longer the case. With an abundance of competitors, PI's who do not deliver data products will not fare well in subsequent peer review. NewDISS must recognize that PI-level data activities are part of broader assemblages of producers and users based within identifiable communities. NewDISS management must use this understanding as an enabling principle for both design and management.

In the future substantially higher volumes of data will be used, resulting from the need for increasingly fine resolution analyses and models, the need to apply multiple data sets from many sources, and the multiple needs of an ever-widening user base. The ESE community of users—from scientists to educators and policymakers—is both broad and diverse. ESE data and information services will have to address the needs of many constituents, within and outside the science community, both nationally and internationally. In addition, scientific investigations will increasingly rely on huge volumes of distributed data and information generation, particularly with cross-sensor or interdisciplinary

applications. It is also important to recognize that while we emphasize the basic science community, the distinction between basic science and other user domains within ESE will increasingly be blurred. For example, applications communities, exemplified in the Regional Earth Science Applications Centers, will increasingly need access to ESE science data and services as pressures to link basic Earth science to national needs continues to grow. An increasing user base should be anticipated. Responding appropriately to these multiple needs should be viewed as a metric for NewDISS success.

Technological change will occur rapidly and the system must be able to take advantage of these changes. Technology is now changing at such a rapid pace that it will be impossible to predict technological solutions even 2 years into the future. Thus, systems and services built with the realization that changing technology presents opportunities as well as challenges will be necessary. The systems and services must anticipate and respond to developments in technology. Modularity offers the best hope of meeting this challenge.

Long-term stewardship and archiving must occur. While networked, distributed enterprise systems offer new challenges and opportunities for information delivery, NewDISS must also ensure a seamless transition to other Government agencies' mandated long-term archive and stewardship roles. Long-term records are central to the analysis and assessment of change. A major focus must be placed on ways in which data are stored and later retrieved for comparative analyses or reprocessing. And, in parallel to flexible and easily changeable systems, there is a requirement for data and information that is readily transferred, and quickly understandable and usable. This report articulates a concept of a NewDISS conceptual framework, which, through adoption of published of "Standards and Practices," allows for such data and information transfer.

NewDISS evolutionary design must move beyond data and information and towards knowledge-based systems. Knowledge represents the relationships that are derived from the information content of data collections, or the relationships between metadata attributes. It is increasingly recognized that knowledge management, as well as information management and data management, provides the advantage of adequately describing the relationship among multiple data collections. Careful consideration of knowledge management issues when choosing and evolving metadata standards will allow interoperability between independent data collections and the orchestrated exploitation of heterogeneous and distributed Earth science collections.

C. THE COMPONENTS OF NEWDISS

The principles outlined in Section B of this document lead logically to a set of NewDISS components based on a dynamic network of interconnected elements, each responsive to its environment, containing capabilities for change over time through feedback with the science community. Structuring NewDISS around science questions and the needs of the science community suggests that the community must be involved directly in its overall design and implementation. This implies a reliance on a distributed and heterogeneous collection of producers and consumers. This data “ecosystem” will necessarily include a variety of interdependent components. As described below, the components of NewDISS have been conceptualized as heterogeneous, including such nodes as “Backbone” distribution centers, PI-managed Mission Data systems, Science Data Centers, and MultiMission Data Systems, and allowing easy participation by scientists and data and services providers.

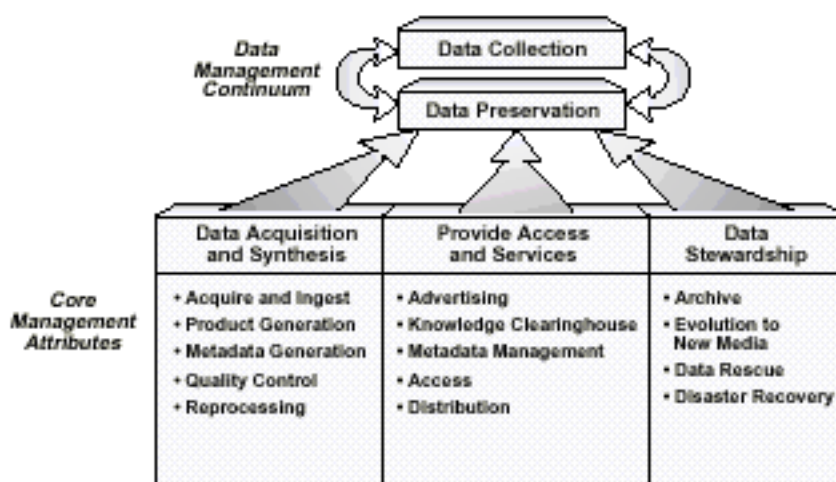


Figure C-1. Concept of functional attributes for Earth science data management

The data and information infrastructure envisioned for NewDISS is obviously not an end in itself. Rather, the components of NewDISS are a means for meeting the objectives for management of NASA’s ESE data. In brief, NASA’s ESE has requirements for collection and synthesis of scientific information, for bringing synthesized data products to bear on unanswered scientific questions, and for preserving data and information for future scientific discovery. Figure C-1 illustrates the free flow of scientific data from active archives to long-term archives and back again, defined here as a “data management continuum” of collection and preservation, and the core attributes on which a successful data management system must be built. The remainder of this section identifies how these needs can be met through the existing and evolving elements of NASA’s Earth science data management system.

C.1 Current EOSDIS Infrastructure and NASA ESE Data and Information-Related Assets

The current EOSDIS system and NASA ESE data and information infrastructure have evolved over time to include a variety of components, which are described briefly below. For

completeness, we have included not only the core science support components, but also activities that support NASA's Office of Earth Science (OES) outreach efforts in applications, commercialization, and education.

DAACs. There are eight DAAC's that serve as primary data depositories for the EOS missions as well as secondary data collectors for EOS products and existing mission data. The DAACs currently employ heritage, heterogeneous systems to generate, archive and disseminate a variety of data including products from NASA satellites and in situ measurements. The DAAC's share data through use of an information and management system. The DAACs have begun receiving versions of the EOSDIS Core System (ECS), which they use to process data from the first series of EOS satellites. In delineating the Working Prototype Federation (described below), NASA has also designated the DAAC's as its "Type-1" Working Prototype Earth Science Information Partners (WP-ESIP's). In the context of the WP-Federation, the Type-1 ESIP's are responsible for standard data and information products whose production, publishing/distribution, and associated user services require considerable emphasis on reliability and disciplined adherence to schedules.

Science Computing Facilities (SCFs). Under the EOS program, NASA's ESE provides funds for SCF's located at the home institutions of EOS principal investigators. In some cases these SCF's are sufficiently robust to allow the EOS principal investigators to generate the higher-level data products for their instruments. In this way, EOSDIS has begun the transition to PI-led mission data processing. In addition to the EOS Instrument team members, NASA provided funds to the Interdisciplinary Science (IDS) Investigators to develop SCF's in their home institutions.

Working Prototype-Federation. In 1998, NASA provided funding to twelve each of "Type 2" and "Type 3" WP-ESIP's for the experimental phase of defining, demonstrating and validating the federation approach to performing several key functions of EOSDIS. The roles and responsibilities of Type 2 and Type 3 ESIP's are defined below.

Type 2 Working Prototype Earth Science Information Partners. Type 2 ESIP's are responsible for data and information products and services in support of Earth system science (other than those provided by the Type-1 ESIP's) that are developmental or research in nature, where emphasis on flexibility and creativity is key to meeting the advancing research needs. In addition, NASA's ESE has awarded technology prototyping funds to the Type 2 ESIP's with the goal of rapidly developing new methods for exchange of environmental information.

Type 3 Working Prototype Earth Science Information Partners. Type 3 ESIP's are responsible for developing practical applications of earth science data for a broader community. Funding for this group is structured so that there is a partnership between NASA and the business or institution that leads each individual Type 3 ESIP. It is expected that ESIPs of Type 3 will become self-sustaining as a result of the mature applications developed by their partnership with NASA.

Mission Data Production, Pathfinder and Other Research Data Efforts. In addition to the data management centers, it is recognized that there are many other team members and activities involved in the ESE data management. For ESE core science activities, these include various research institutions funded under the NASA ESE Research and Analysis budget.

In addition to data and information production and distribution efforts driven by ESE's science research activities, ESE supports a variety of information elements that contribute to its applications, commercialization, and outreach efforts. With this in mind, the authors of this report have endeavored to construct a framework that is easily extensible to research and applications activities. Some examples of ESE applications activities include, but are by no means limited to, the Type 3 Working Prototype Earth Science Information Partners, The Affiliated Research Centers, and the Regional Applications Centers.

C.2 Proposed Near-Term NewDISS Components

As previously stated, NewDISS will consist of highly distributed, heterogeneous components. These widely varying participants will be responsible for executing the data management functions described in section C.1. Because the NASA ESE already has made considerable investment in data activities, it is imperative that the near-term NewDISS organization components should leverage the structure of existing systems and services. The future NewDISS structure could be quite different, however, as existing data center activities take advantage of technological innovations and otherwise evolve to meet future science-driven demands for data management.

Figure C-2 conceptually illustrates NewDISS data flow. NewDISS provides a means for opening numerous new channels for Earth science satellite (and other observational) data streams to reach the user community. Such channels will flow to users both directly from mission data systems and also via many intermediate information providers. This approach to Earth science data flow permits comprehensive intellectual exploitation of the data.

While we leave the development of a NewDISS architecture or architectures to subsequent studies, the following is a concept of the various institutional components of NewDISS for the next 6-10 years.

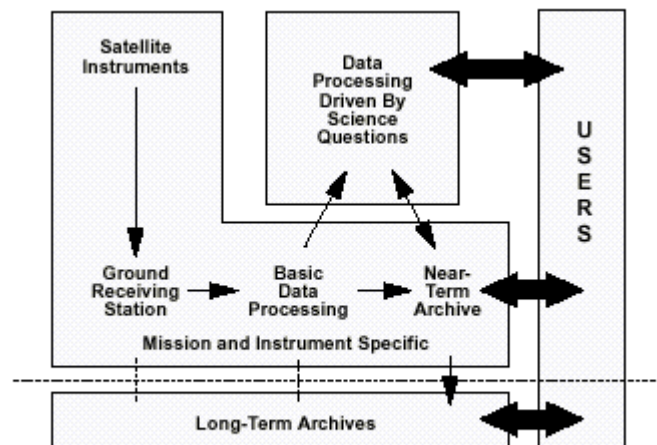


Figure C-2. Conceptual NewDISS data flow

Backbone Data Centers. These centers, most likely evolving from some of the current DAAC's, will address NASA's responsibility for preserving and protecting the large volumes of data from the ESE satellite missions. One of the primary roles of the backbone data centers will be to preserve the basic data. Clearly, NASA can provide a considerable amount of existing infrastructure and technical skill needed to provide satellite mission data downlink and "level 0" or "level 1" data processing. Teaming NASA missions with Backbone Data Centers in the Announcement of Opportunity (AO) process for backup or for generation of basic data products may well be an attractive option for handling some of the core data management requirements of NewDISS. Another role for the Backbone Data Centers will be to acquire products agreed to be scientifically important for preservation and to prepare all these data for long-term archiving. These data centers, staffed by professional data managers, provide historical experience and proven capabilities. As such, they provide a means for risk mitigation against the failure of one or more of the NewDISS components by serving as backup centers for the other parts of the NewDISS. These data centers would most likely be few in number to ensure the cost-effectiveness of the NewDISS. The Backbone Data Centers will increasingly need to support access to level 1 data by PI-led data analysis projects. This means the centers must provide mechanisms for discovery of data and rapid, automated access to data.

Mission Data Systems. These data systems are specifically affiliated with instruments or satellite systems. They are either PI-led or facility/project-led. They provide key measurements and standard products from NASA-supported satellite instruments. The key characteristic of the mission data systems is that they will be proposed, engineered and implemented as part of an ESE mission. It is anticipated that these Mission Data Systems could leverage the activity at the current ESE data management infrastructure: the ECS flight operations and science data systems and the other hardware and software infrastructure at the DAAC's, the ESIP's, and the SCF's. Mission Data Systems will be responsible for their data management functions during an Earth-observation space flight mission. These data systems will be funded by the mission selectee through the ESE flight programs and will be selected by competitive selection for future ESE missions.

Science Data Centers. There is a recognized emerging need for Science Data Centers. These data centers will collect data from multiple missions for a user community focused on a single research question. There are several examples of these types of Science Data Centers in NASA's Space Science Enterprise. These centers are targeted at specific science questions (perhaps from the NRC Pathways Report) and/or science disciplines, and they directly support research and data analysis for specific research questions. In many cases, these Science Data Centers may evolve from one or more of the current DAAC's (for example, NSIDC or SEDAC) from the ESIP-2 community (for example, a Land Cover/Land Use Change data center), or from the EOS Interdisciplinary Science Investigations. It is anticipated that these data centers will be selected through competitive peer review.

MultiMission Data Systems. A third type of data center is the MultiMission (or Measurement) Data Center. The type of data activity to be carried out by such a data center is the generation of consistent time-series geophysical parameters, an activity exemplified by the National Oceanic and Atmospheric Administration (NOAA)/NASA Pathfinder Datasets

program, which was funded by NASA's ESE and carried out by PIs at various institutions. These efforts will take on more importance in the future, since NASA ESE has the requirement for generating time-series of geophysical parameters, while the EOS mission strategy has evolved so that it is now designed to accommodate technological change. Thus, these efforts will include construction of the long-time scale datasets from more than one NASA (or other) mission.

Infrastructure Components. In addition to the data center components listed above, the authors of this report recognize that there are other "infrastructure" or "glue" pieces necessary for a complete NewDISS. It is recommended that the NewDISS infrastructure include active liaison with service providers both within NASA and within the private sector for procurement of common operations activities in order to move to more effective operations. ESE liaison in this regard should ensure end-user feedback to the infrastructure service providers, and should also ensure that Mission teams have the correct information to make choices about whether to use NASA or alternative service providers. Three infrastructure elements requiring NewDISS liaison are highlighted below.

Mission Operations. All of the Mission Data Systems will need to address satellite/instrument command and control and data downlink. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives, such as commercially provided or university support services.

Networks. All of the data centers will need to address connectivity issues as part of their ongoing activities. Again, selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives.

ESE Long-Term Archives. The long-term archive of ESE data is not addressed in the NewDISS report, since archival in perpetuity is not the purview of NASA for Earth science data. NASA has signed Memoranda of Understanding with USGS for long-term archival of land processes data and with NOAA for oceanographic and atmospheric data, and currently works with these two agencies to coordinate implementation. Elsewhere in this report we do mention NASA's role in preparing data for transfer of responsibility. Recently, NASA cosponsored a workshop on Global Change Science Requirements for Long-Term Archiving (USGCRP 1999), which provides guidance on long-term archive essential programmatic functions and characteristics to inform such data preparation and responsibility transfer.

Hierarchy of Services. Easily accessible tools that enable the user community to better discover, access, understand and use Earth science data and information are increasingly being broadly provided. The NewDISS era will enjoy the availability of a spectrum of services, including ever more sophisticated tools. The integration of data with services can be accomplished by the emergence and wide deployment of very simple open standards. Publication of service catalogues will provide wide exposure of and enable intelligent selection of available services.

The paragraphs above identified a variety of interdependent components of a NewDISS, including Backbone Data Centers, Mission Data Systems, Science Data Centers, and others.

A key goal of NewDISS is to harmonize and aggregate these various, disparate, and perhaps numerous elements. However, the management of these elements is not formulaic. The potential connections in a fully integrated and comprehensive data service are many, and there is no well-defined process to prioritize which connections are most important. It is impossible to anticipate the numerous contributions of data providers or the needs of customer groups, or how the customer groups will access and apply data.

Although we conclude that it is impossible to establish a formulaic specification for the totality of NewDISS, it is still our conclusion that the problem can be broken down into manageable subunits. The approach to a NewDISS would then be to specify the behavior of NewDISS components and the key interfaces among them, without necessarily specifying how these components would be implemented.

One means for implementing the NewDISS approach is to reduce risk by compartmentalization: minimizing risk in the crucial NewDISS system interfaces, but allowing an acceptable level of risk in those components of the system that address a rapidly changing user or technological environment. NASA's ESE will also need to achieve balance in the manner in which Earth science data management components and interfaces are defined and implemented. In the NewDISS era, NASA's ESE will increasingly play the role of defining the "rules of engagement" between and among NewDISS system elements, rather than adhering to the traditional role as either developer or procurer of the system.

The key objective in this approach is to avoid specifying the implementation of NewDISS components such that new services can be adopted if they already exist or if they can be easily developed from community norms. This does not mean that NASA/ESE will never participate actively in defining NewDISS components or interfaces. For some tasks, NASA/ESE may select and fund specific activities to work within the selected rules. For example, a specific satellite-observing mission may be funded by NASA to produce certain critical data products. However, NASA will not specify the details of the institution that implements this activity; instead, NASA will only ensure that the rules are followed. The mechanisms for achieving this end are elaborated in the Sections D and E of this report.

C.3 An Implementation Scenario

The implementation of NewDISS depends of course on how the principles of this report are applied to the various institutional components, standards and practices sketched in the previous sections of this report. It is important to note, however, that these principles and practices cannot be implemented uniformly across all NewDISS components. Spaceflight Mission Data Systems must necessarily adopt a different behavior than that adopted by Science Data Centers. Although it is beyond the scope of this report to provide a complete definition of how these principles will ultimately be applied in the actual implementation of NewDISS, the following scenario provides additional explanation on the behavior and interfaces of NewDISS elements.

Science Discipline Data Systems. For science data systems, including those at the laboratories (or on the desktops) of individual investigators, data are acquired and analyzed

in the context of specific scientific disciplines: climatology, hydrology, oceanography, vulcanology, etc. The standards and practices governing the acquisition, archiving, documentation, distribution, and analysis of these data are, de facto, those established by these discipline-specific scientific peer groups. An ESE NewDISS must recognize and embrace this tapestry of communities and disciplines, and their discipline-specific standards and practices.

In the case of Science Data Centers and MultiMission Data Systems, NewDISS participants will affiliate with one or more disciplinary communities, as determined by their scientific (PI's) or service (data centers) orientation. These communities will have established their own standards and practices, and will be responsible for exposing a uniform set of interfaces (formats, services, protocols, etc.) to one another and to other components of NewDISS. It is important to note that these interfaces are established by community consensus, and may complement other interfaces internal to the community. What matters is that the community agrees on a "common face" for its participation in ESE NewDISS, and that other elements of NewDISS are able to accommodate these standards and practices when necessary.

In addition to accepting existing discipline standards and practices, NewDISS will have to accommodate the reality that these disciplinary communities will evolve over time. The traditional disciplines will grow or shrink with the number and productivity of their participants. Additionally, new disciplinary communities will emerge with their own suite of participants and practices.

Mission and Backbone Data Centers. Backbone Data Centers and Mission Data Systems will play a dual role in establishing and adopting NewDISS interface standards. On the one hand, they may establish new—or extend existing—interface standards to meet the needs of specific space flight missions. On the other hand, these NewDISS elements will need to support the interoperability, openness and survivability of interfaces defined by specific research communities.

One of the obvious characteristics of space flight missions is that they must adhere to a specific launch schedule. Consequently, all launch-critical components and interfaces must be designed, implemented, tested and certified before launch. The constraints of a launch schedule may be such that it is impossible to wait for community consensus for interface standards. In such cases, Mission and Backbone Data Systems will need to design and adopt interface standards that may not be compliant with community norms. Such a consequence may seem regrettable, but it may also be unavoidable. In such cases, the responsible NewDISS elements will need to publish and maintain interface definitions, and they may be called on to ease the acceptance of such interfaces with the affected scientific communities.

Backbone Data Centers and Mission Data Systems will provide an institutional base for many of the NewDISS professional data managers. As defined in the previous section of this report, NewDISS Backbone Data Centers and Mission Data Systems may often be called on to play a role as curators of science data and data interface standards, and may even be called on for operational science product generation. Certainly these institutions, and the personnel that make them up, are expected to be on the front line of user services—and must be expert

in the myriad issues related to this task, from answering technical questions about data products to making restitution for lost orders. In all cases, Mission and Back-bone Data Centers will need to support the interoperability, openness, and survivability of relevant interfaces defined by the science communities.

D. NEWDISS INTERFACES, STANDARDS AND PRACTICES

There is evidence from the WP-ESIP federation “experiment” to suggest that a new model is emerging which provides insight for modeling NASA data systems. This model has been called an information economy. In this model, which we can refer to as a *Domain Brokering Model* (DBM), data facilities centered on particular domains of expertise, or *Domain Brokers* (DBs), work in tandem with specific associated Backbone Centers, or DAACS as they exist today. The brokering relationship focuses on enhancing the usefulness of the data and information for particular user communities.

What follows is a concept for NewDISS that considers the role of central, Backbone Data Centers in collaboration with science, or domain-specific, data centers. Using the terminology of the current EOSDIS configuration, the DAACs have responsibilities for managing large archives of platform or sensor-specific data. While DAACs have skillfully demonstrated the ability to manage vast archives of data, most domains and user communities require more than data and data products: they need targeted information, access to methods and techniques for analysis, specialized products, and technical, scientific or applications assistance in ordering and using NASA data. What emerges from this understanding is the concept of a suite of Data Brokers, each acting as an adjunct to the existing DAAC structure. In essence, Data Brokers facilitate the wholesale-to-retail end of the business in this information economy.

At present, the relationship building between data centers and domain brokers is not explicit. NewDISS makes this relationship explicit, through the setting of core rules and standards, described below, including various core rules of engagement. This does not, however, imply that NASA needs to fund or support all necessary domain brokers, although it may be prudent to establish some critical ones for specific areas of science priority or emphasis.

The Data Center/Data Broker model supports the notion that jointly entities such as ESIPs, acting as data brokers with the DAACs, provide better service to a larger and more diverse customer base than with either alone. It also provides a mechanism for evolution of the DAAC system through targeted and domain-specific brokering of both data products and services, directed on the one hand by agency and program priorities, and by user needs on the other.

In the sociology of the Data Center/Data Broker framework, Backbone Centers would functionally be the locus for source data, products and core standards, as well as long-term stewardship. Data brokers could take several forms according to evolving community standards and practices, including for example:

- *service providers*, which provide workbenches or tools for subsetting, format conversion, search and retrieval, catalog interoperability, client-side interfaces and so forth,
- *domain portals*, which provide a data portal to the DAACS, brokering for data orders, custodial and re-distribution services for data which have been ordered, (e) provide a

clearinghouse or cooperative function for sharing among community members, value added data products: (e.g. atmospheric correction) or value-added derived products (e.g. digital GIS layers).

NewDISS needs to institute management and operational capabilities that enable the development of a Data Center/Domain Broker model. Such capabilities need to focus on characterizing what component services data brokers must have, and how the relationships with between DAACs and Data Brokers can best be developed within established rules or governance structures. the emergence of open standards such as those being promulgated by groups such as the Open GIS Consortium and others, the information economy of producers, consumers, and service providers can provide rapid response to changing technological capabilities and science priorities. This view emphasizes less the architectural specification of the components themselves, and more the *relationships*, or modes of interaction, between them. The NewDISS economic relationships among data producers and consumers allows for growth and expansion of the number and diversity of parts through a set of rules, which are termed here "core and community standards."

D.1 Core and Community Standards

There are a few key concepts to consider in understanding how the basic building blocks (institutions and functions) of the Earth science collection-to-publication process are organized by NewDISS partners. First, any or all of these functions may be performed by multiple, autonomous entities. Therefore, functional interfaces between each of these building blocks must exist. Second, as these basic functions are organized by investigators and institutions, NASA cares most about what they do (the services provided by the individual and institutional partners) but considerably less about how they do it (the specifics of how the NewDISS partner internally connects its unique set of functions). In other terms, there are core and community attributes to NewDISS. These concepts are defined below.

Core: That which NASA has a vested interest in controlling

- Legal responsibilities
- Long-term responsibilities
- Willing and able to impose on all participants
- Characterized by "continuity"

Community: That which NASA has a vested interest in NOT controlling

- Characterized by subsidiarity: Left to one or more appropriate communities
- Characterized by evolvability: Multiplicity promotes innovation
- Characterized by practicality: There is a limit to what NASA can manage and afford
- Characterized by "growth"

NASA is relatively unconcerned with the specifics of how NewDISS partners (both investigators and institutions) internally implement data collection, storage, access, retrieval, analysis, and publication functions or how they are "wired up" internally at a partner's site.

NASA does, however, care that the functional interfaces of one partner are openly and publicly defined to another.

Previous discussion in this section focused on interaction among NewDISS investigators and institutional partners. However, NewDISS cannot simply define the standards and interfaces for data exchange among its partners; there must also be a definition of how partners can aggregate their products and services in support of various user communities. NewDISS partners may aggregate into NewDISS clusters (also synonymously referred to as brokers, bundles, or portals) and such aggregation may be motivated by any of a number of factors, for example: affinity by discipline, programmatic, nationality, et cetera. From the NewDISS perspective, a community of partners is defined both by the specific implementation of interface standards within the group and by the methods by which the group as a whole exposes its functional interfaces to other groups, individuals and institutions. Simply stated, the standards do not define the community, the community defines the standards. On the other hand, community and individual decisions on implementation of standards continue to exist within the context of NewDISS standard practices (described in Section D.2), and within the limits of the process for choosing standards (described in Section D.3).

D.2. Core Standards and Practices

The following are standards and practices that must be followed by all NewDISS partners when interchanges of data and/or services across interfaces are required. Communities may follow their own standards and practices; however, the core standards and practices must be adhered to for inter-community interfaces.

- **Long-Term Understandability:** Data shall have sufficient information associated with them to allow a user to read and use them. It is especially important to capture all the information about the origin of the data, history, algorithms, antecedent datasets, versions, etc. to permit the use of data for decadal scale climate research. Researchers decades into the future shall be able to understand the documentation associated with the data. For example, the encoding standard used to structure the data must be specified..
- **Longevity:** Any data that are intended for long-term use shall be adequately preserved and maintained, including periodically refreshing the storage media, updating the technology and ensuring that the data remain readable.
- **Open Access to Data:** The data developed by NewDISS partners under government sponsorship shall be openly accessible. Any copyrighted data that traverses an inter-partner interface shall become openly accessible. The interfaces between NewDISS entities shall be open (non-proprietary). Waivers may be granted in special cases.
- **Publicly Documented Formats:** Non-proprietary data that are exchanged among NewDISS partners through the interfaces shall be a one of a small number of agreed upon formats that are publicly documented.
- **Persistent Identifiers:** Any dataset that needs to be addressed independently shall have a persistent, unique identifier (analogous to a URN). Such identifiers shall be permanent throughout the life of the dataset. In fact, it is desirable that the identifiers never be reused even if some of the datasets are retired. The persistent identifier must be independent of

the storage system used to hold the data., implying the ability to move data between federated systems.

- **Searchability:** As a minimum, users shall be able to search for data by specifying temporal and spatial selection criteria. The interfaces to the catalogs will be published to enable federation across catalogs.
- **Charging:** No NewDISS interface shall require nor preclude the imposition of a charge for use of the interface. It should be possible to charge for data or services while it should also be possible to provide them free of charge. That is, the *interfaces* shall be policy-neutral.
- **Performance:** Performance specifications, including information transfer through an interface, shall be agreed-upon and documented. These shall be negotiated between partners

D.3 NewDISS Core Interface Selection & Maintenance Process

The set of NewDISS Core Interfaces must evolve over time to take advantage of new technologies and new ideas. For example, an information system specified 10 years ago would not be web based because the web did not exist then. Certainly many information systems used today have web-based access. How will NewDISS select and evolve its Core Interfaces? NewDISS Core Interfaces are expected to be openly published, public interfaces, similar to the interface standards released by standards bodies. How do successful standards bodies select and evolve its interface standards?

Most standards bodies use a set of defined processes to select, maintain, and evolve their interface standards. Different standards bodies have had varying success in promulgating widespread use of their standards.

Two examples of very successful standards groups using processes to determine the selection of interfaces are shown below. The first example is the Open GIS Consortium, a leading standards body for interoperable geospatial information systems. The second example is of the W3C, which defined the various versions of the html standard, one of the most widely used application standards today.

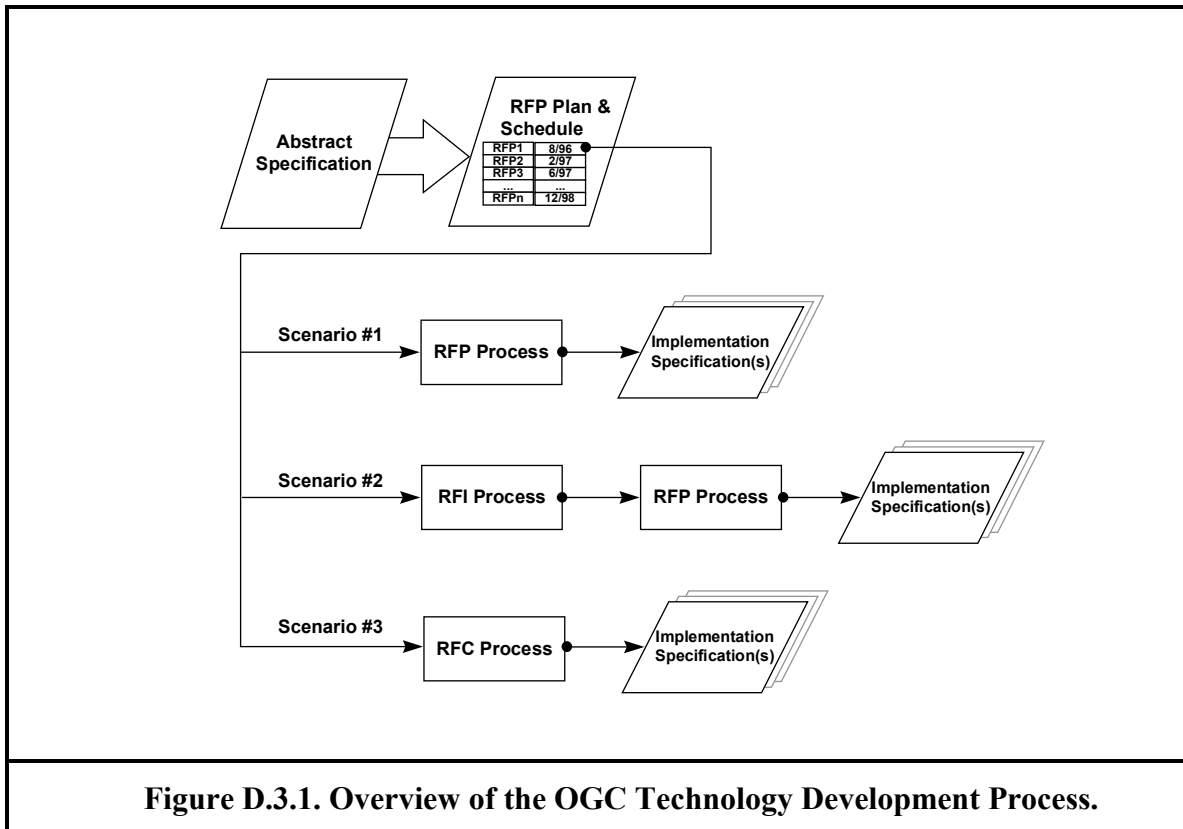


Figure D.3.1. Overview of the OGC Technology Development Process.

Figure D.3.1 shows the overview of the Open GIS Consortium (OGC) Technology Development process. The basic goal is to generate an implementation specification from an abstract specification. The abstract specification contains the requirements for the implementation specification. The Request for Proposal (RFP) Plan and Schedule contains the technical content for a series of RFPs and a schedule for obtaining Implementation Specifications for that content. In Scenario 1, the OGC staff issues the RFP, multi member teams respond to the RFP with a proposed Implementation Specification. This process results in at least one implementation specification that can be adopted. In Scenario 2, an RFI process is initiated before the RFP process is started. This also results in at least one specification that can be selected. In both scenarios, the technical team (in this case the OGC staff) generates the RFP and the vendors generate the response to the RFP, which is the implementation specification. In Scenario 3, a team of vendors initiates an RFC (Request for Comment) on an implementation specification. This is not a response to an RFP from the technical team but an unsolicited implementation specification that can then be adopted. The OGC Catalog Implementation standard was the result of a response to an RFP while the World-Wide Web (WWW) Mapping Testbed Mapserver Implementation Standard is the result of an RFC brought by the participants of the WWW Mapping Testbed after the initial testbed exercise.

The second example of a standards group process is the process employed by the World-Wide Web Consortium (W3C), which brought http, html, sgml, and vml standards to the

world. The W3C process to produce implementation standards consists of a linear progression of 4 stages a document must pass through in order to become a W3C approved standard. A document is a "Working Draft" when it represents work in progress and a commitment by W3C to pursue work in a particular area. A stable working draft becomes a "Candidate Recommendation" when the Director has proposed to the community for implementation experience and feedback. A "Proposed Recommendation" is a Candidate Recommendation that has benefited from implementation experience and has been sent to the Advisory Committee for review. A "Recommendation" reflects consensus within W3C, as represented by the Director's approval. In each stage of the W3C process, there is no guarantee of the document advancing to the next stage. Some of the documents will be dropped as active work and some will be published as "Notes."

NewDISS will use a process-driven approach to selecting, maintaining, and evolving its Core Interfaces because it is clearly impossible to define the specific implementation of standards for all possible interfaces in the functional and aggregation layers of NewDISS. Furthermore, it is not even desirable to do this since the "evolvability" of NewDISS relies on the ability to change the implementation of a standard (even as the need for an operational interface remains constant). Just as clear, however, it is the need for decision making, adoption of standards, and rigid adherence to standards in core ESE functions such as satellite data capture and low-level data processing. Such adherence to core NewDISS standards may apply to other areas as well. Figure C.5.2 illustrates that a process must exist by which the standards for implementing NewDISS functional interfaces are allocated to continuum for NASA management.

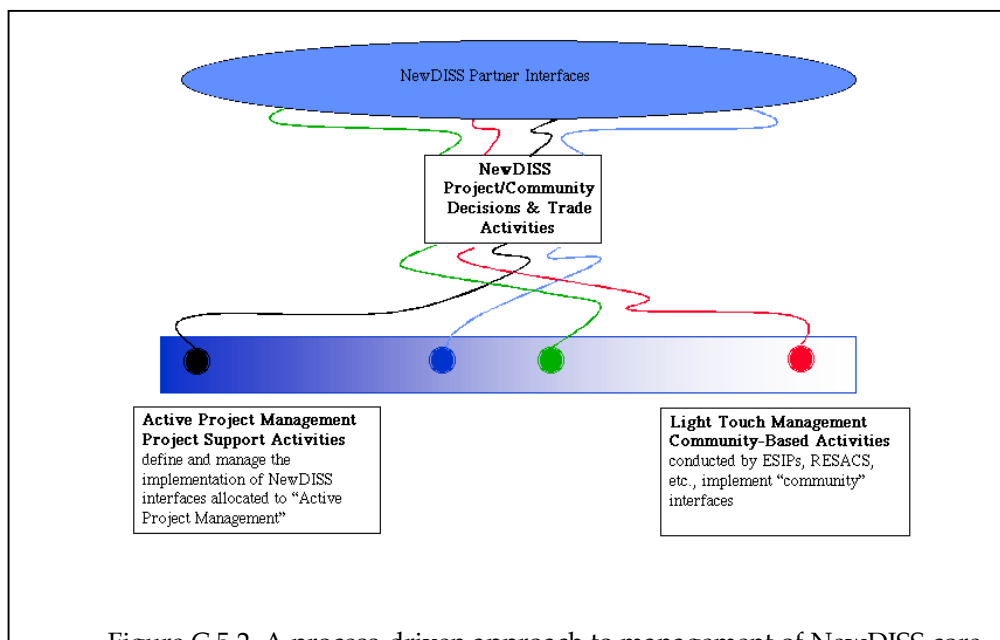


Figure C.5.2. A process-driven approach to management of NewDISS core and community interface standards.

E. MANAGEMENT AND LEADERSHIP

This section is concerned with effective management—not of data but of the institutions, resources, standards and practices that are needed to achieve the functional goals of NewDISS. Previous sections of this document made the case that an adequate solution exists to the challenges of Earth science data management. Furthermore, it was asserted that such a solution can be found (and optimized over time) in a flexible framework that integrates the numerous, disparate elements of NewDISS: various data service capabilities, common interfaces, and computing and communication capabilities both from within and outside of NASA. Since it is our position that the elements for an adequate solution to NewDISS exist, then the most formidable task facing the development of a successful NewDISS is truly the management of these elements. As NASA moves toward more distributed, heterogeneous data service capabilities many of these NewDISS elements will be designed, developed and operated by disparate entities, such as by individual or clusters of investigators, Mission Data Systems, partner data centers and others. Thus, a viable management function may be the single most important deliverable that NASA can provide. The sections below lay out NASA's role in delivering NewDISS management, focusing on three key aspects: system diversity and integration, governance structures, and metrics.

Before beginning the discussion of management, it is necessary to address NewDISS “leadership,” and to define the role that NASA and its advisors must play in conducting the tenor and tempo necessary for NewDISS success. The traditional NASA agency program/project management approach used in past missions did not support an adequate leadership function. Traditional NASA management assumed that a well-defined set of requirements could be generated and that a contractor could be engaged to implement those requirements. In fact, however, the specifications for complex data systems are neither simple, static, nor specific, and in many cases solutions do not exist for some of the requirements. Clearly, there must be a leadership function that takes ownership of both the changing requirements and the functionality of the NewDISS.

Leadership must identify requirements, set priorities, and must also link requirements to cost and functionality. This includes continued re-evaluation of the requirements in the face of changing programmatic, scientific, and technological environments. The underlying philosophy in NewDISS leadership should be to ensure that it is very capable of organization and interorganizational change. The leadership function must be provided with recognized authority to lead the component organizations of the effort. The relation between leadership and flow of money must be clarified and aligned to provide the basis for an incentive structure to link costs, requirements, and functionality.

The management structure and process in NewDISS must support the flexibility required for effective leadership. Ultimately, *leadership* provides the vital, thinking flexibility needed to appropriately respond to changing external factors. In contrast, *management* addresses the implementation and execution of the tasks needed to provide a viable data service. Management without leadership will not be successful.

E.1 Managing Diversity and Integration

Diversity in any system imparts a certain degree of resilience: the ability to meet and adapt to a changing environment. Equally important, NewDISS diversity can be a hedge against changes in the user community and rapidly changing information technologies.

We recommend that NASA support a spectrum of heterogeneous participants and technological approaches to NewDISS. Such diversity should especially be considered in the investigators, organizations and institutions that make up NewDISS. Section C of this report describes the infrastructure and functional components of NewDISS. The responsibility for the actual implementation of these components may be given to Principal Investigators at universities or other private institutions, or to Project Managers at government-funded facilities or research centers. In allocating responsibility for the implementation of NewDISS, we have recommended (see section B) that NASA employ competition and peer review in the processes used for choosing NewDISS components. Furthermore, we recommend that NASA empower science investigators with an appropriate degree of responsibility and authority for NewDISS data system development, processing, archiving and distribution. This will help ensure that NewDISS elements are tied to the science, and that the system developed for each experiment or mission matches the needs of that mission.

NewDISS management must concentrate on integrating suitable existing data service capabilities, while also identifying and providing a means for integrating capabilities that do not yet exist. This is a much more abstract function than current project management, and represents a significant cultural shift for current management style. NewDISS managers may integrate services and technologies from a variety of sources, for example, NewDISS should encourage the development of appropriate new information management technologies from other groups within NASA or other parts of the government. NewDISS should also be open to the infusion of new technologies developed by industries that a few years ago were completely unassociated with digital information management, but are now leaders in the field, e.g., the banking, entertainment and retail industries. Finally, NewDISS should not be afraid to adopt accepted, orthodox data management solutions where appropriate.

E.2 Structures for Management and Leadership

Because NewDISS will consist of a heterogeneous mix of components, participants and services, the NewDISS management structure naturally lends itself to some form of shared governance. The goal of such a management structure is to bring competitive and diverse market forces to the data enterprise. In this shared governance, NASA would have both a direct management role and a role as an equal partner. In its direct management role, NASA would generate AO's, elicit peer reviews, evaluate proposals and monitor the performance and progress of grants, contracts, and cooperative agreements. NASA would also lead the determination of data policy issues, copyright, and liability. As an equal partner, NASA would work with others in key issues such as standards and technology infusion.

Naturally, a broad and diverse aggregation of Earth science communities exists outside of the current ESE federation experiment, and those communities have managed data for a long time. Without any outside management at all (or perhaps with a minimum of fuss) various disciplinary and interdisciplinary communities and clusters of such communities have centered on specific scientific problems and data management issues. There already exists agreement at the community level on standards and practices, formats, and tools. NewDISS should be aware that these communities exist, and should use them as focal points for management of data and services. Recognizing that individual PI-based data activities are part of broader assemblages of producers and users based within identifiable communities would be an enabling principle for both design and management.

NASA ESE is currently engaged in an ESIP Federation Experiment, from which we may draw some early lessons. The ESIP Federation can point the way, as NewDISS is itself a federated model. The ESIP Federation has shown that, if an environment is created that allows for evolution, such a framework will be able to meet NASA's growing needs for integrating data from multiple sources, including sources beyond its control (e.g. from other agencies).

Initially, as a response to addressing the issue of system interoperability, the Federation proceeded by organizing itself into "clusters," or groups of investigators that work together to provide specific services to a targeted user community. At a grass-roots level, the ESIP Federation is integrating community interfaces and standards, resulting in the availability of more versatile tools, which are useful and easy to use by the intended communities. The ESIP Federation has also chosen the capabilities to be included in Version 1 of its System Wide Interoperability Layer. These activities have been accomplished at a prototyping level through the relatively small annual investments provided as part of the original Federation Experiment process for ESIP data center integration.

The ESIP Federation has expanded beyond the NASA prototype stage, and includes membership of additional USGCRP agencies, and more applications, commercialization, and education organizations. This expansion of the ESIP Federation broadens the reach of ESE products in science and applications areas.

NASA recognizes, as does the Federation itself, that the ESIP Federation must be evaluated – in the final analysis, the customers and the community will assess how successful the ESIP Federation is. The perceived benefits are at least twofold: the elimination of existing programmatic "stovepipes", and the capability to organize a more distributed set of heterogeneous systems and services than are now provided by ESE's large, centrally-managed EOSDIS.

We recommend that lessons learned from the current, experimental ESE federation should be used as a step towards the NewDISS, and that the Federation Experiment proceeds with this evolution in mind.

In establishing NewDISS governance structures, serious consideration should be given to placing some management elements outside of NASA. While there are no robust, successful models for how to manage complex data systems, the possibility that an external organization, a nonprofit entity perhaps, whose very success is dependent on the delivery of satisfactory data service, should be considered. Private-sector data management initiatives also exist, and should be considered as part of the collection of tools available for NewDISS management. As another alternative, a substantially altered NASA management structure, with a bias toward equal partnership, rather than centralized management, could be considered.

Finally, it is safe to assert that the scientific user community, the first-line customers of the data service, is skeptical of any centralized management by NASA. We have already explained ways in which centralized management or development of NewDISS is inappropriate. Given the complexity of the task, the volatility of the technology, and the changes in scientific priorities, a centralized approach is too inflexible and increases the risk that large portions of the data system will be vulnerable to single-point failures. However, ultimately, if Earth science observations are going to be used in policy decisions, then the integrity of the observations must be documented and preserved. The need for rigorous data management and long-term stewardship suggests that certain aspects of central control and management integration are desirable. There are areas where firm authority is needed. However, as long as discovery of how to solve a problem remains an important ingredient of the solution, centralized control is inappropriate. We recommend that the NewDISS transition plan (see section E.6) should address the important question about precisely which functions of NewDISS must be centralized and which must not. In making these assessments, the transition team may wish to review the TRMM Science Data and Information System (TSDIS), the Lightning Imaging Sensor (LIS), and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) experiences that are briefly sketched in the section of this report on cost considerations.

E.3 NewDISS Partnerships

Next generation data systems and services will increasingly engage partners outside of NASA's ESE community. Several of these key partnerships are highlighted below.

Partnerships Across the Global Change Research Program. NewDISS will need to facilitate partnerships with other agencies within the Federal Government, such as NOAA, USGS, EPA, etc., in the context of global change research, and the USGCRP. Stronger linkages would leverage resources found in other agency efforts and could reduce redundancy and foster synergy across the U.S. Government.

International Partnerships. NewDISS data systems and services cannot focus solely on national programs. A close working relationship with the international global change research community can be an effective way to gather science requirements and prioritize data products. It can also be an effective way to ensure planned data assets and services support a wide community of users. Moreover, external data initiatives within international partnerships can efficiently complement and enhance the ESE efforts. Participation by scientists in the International Geosphere-Biosphere Programme (IGBP),

International Human Dimensions Program (IHDP) and World Climate Research Program (WCRP) has proven to be effective in mobilizing the international research and monitoring community. Similarly, Data and Information Systems Framework Project of the IGBP (IGBP-DIS) has provided a useful mechanism for launching new international data initiatives in support of the science community. The utility of the NewDISS will, in part, be a reflection of its success delivering useful information for major international programs related to climate change and global change policy. The activities of the Intergovernmental Panel on Climate Change and the Framework Convention on Climate Change are current examples of important outlets and user communities within the international arena. These are important, science-driven fora where the NewDISS will prove its success or failure.

Multilateral Linkages With Other Space Agencies and Observing Programs. There can be considerable efficiency in developing cooperative links to the information systems and services projects within the various foreign space agencies. NewDISS should recognize the efforts going on elsewhere, in particular the various efforts undertaken by CEOS, which will provide a forum for developing collaborative agreements with other space agencies. International global observing programs are developing the Global Observing Strategy (GOS) with the goal of integration and fusion of datasets from a variety of sensors, including those flown by the international community. NewDISS must develop active linkages to these on-going and evolving international information systems. These programs provide an important link to the nonspace-based observations and data, which will be an important part of the NewDISS systems.

Linkages With the Private Sector. The private sector will continue to be an important ESE partner. First, many private sector concerns represent opportunities for cost-effective development of value-added products aimed at targeted audiences and stakeholders. Second, private Earth observation missions are important sources of data for the global change research community. This reciprocal relationship between ESE NewDISS and activities in the private sector should be emphasized and capitalized upon. Indeed, a distributed NewDISS may be the best way to identify how private sector components can “plug in” to the system. Joint strategic partnerships that focus on technology problems or science-application interface problems can be to the advantage of both communities. Moreover, it is important to point out that information systems are now being planned and developed within the commercial sector. On the one hand, this will place added need to ensure compatibility between NewDISS and the private sector systems, particularly in terms of metadata formats and protocols. On the other hand, close partnering will ensure an efficient means for managing the changing landscape of technological change.

Information Technology Partnerships. A flexible, distributed NewDISS must plan for, and adapt to, innovations in information technology. NewDISS managers and participants must identify technology gaps that are essential to meeting its mission. They must be responsible for formulating and prioritizing technology needs by identifying and planning for projected information technologies needed to support the long-term objectives of NewDISS. NewDISS must also establish partnerships with prototype developers, both

within and outside the Government, to ensure that prototypes are properly targeted toward the NewDISS working environments.

Partnerships in Networking. The growth of data and telecommunication networks—capacity, speed, access—has been remarkable. In only the last 4 years, the use of the Internet has grown from 5 million hosts to 40 million hosts. But most of this growth, and most of the network capacity and developments in networking technology (upon which NewDISS will rely) will occur outside of NASA ESE purview. It will be increasingly important that NASA establish focused, objective-driven partnerships with university, government, and the private sector initiatives, including the Internet-2 consortium, Abilene, the Next Generation Internet (NGI), and other high-performance network initiatives. Thus, all the pieces of NewDISS will not, and probably should not, be under direct NASA management or control. But collaborative partnerships for applications prototyping, testbeds, and standard setting should be a way to ensure ESE needs are met.

E.4 Cost Considerations

The NewDISS team recognized that a detailed estimate of the costs for NewDISS is beyond the scope of this report. Should the Earth Science Enterprise decide to implement the NewDISS model, an estimate of the funds required will be an important task for an implementation team. However, some rough guidelines can be culled from a recent NASA Science Information and Services (SIS) study, and from a survey of cost-efficient ESE Mission Data Systems.

The SIS study looked at NASA science missions across the agency and separated the costs into four categories: Mission Research and Development, Mission Operations, Science Information and Services, and Science Research. The allocations from recent NASA missions are shown in table D-1 below.

Table E-1.

<i>NASA Enterprise</i>	<i>Mission R&D</i>	<i>Mission Operations</i>	<i>Science Information and Services</i>	<i>Science Research</i>
Space Science	64%	8%	7%	21%
Earth Science	54%	9%	14%	23%
Microgravity	75%	9%	1%	15%
NASA Average	61%	9%	9%	21%

The SIS study showed that Earth Science spends almost twice as much as a percentage on SIS as does Space Science. A number of factors contribute to this difference. Earth Science missions as a rule produce orders-of-magnitude more data than do Space Science Missions. This is due, in part, to earth orbiting missions' close proximity to the ground receiving stations as well as the number of spectral bands and the high resolution of Earth sensing instruments. Because of the enormous data volume, the Earth Science data systems are larger. Another contributing factor to the higher percent-age cost of Earth

Science data centers is the Earth science user community, which is larger and more diverse than the Space Science community. Space Science mission data are distributed to a relatively small number of discipline science users. By contrast, Earth Science data are used by a very large group of scientists (approaching 30,000) in many different disciplines. In addition, NASA Earth Science data are distributed to, and used by, an increasingly large number of applications users.

A quick survey of NASA data centers provided several examples of very cost-efficient Mission Data Systems, like the TSDIS, and those for the LIS and the SeaWiFS. It is interesting to point out that in these cases the TSDIS, LIS and SeaWiFS Mission Data Systems did not have any responsibility for data archive, for distribution to the general user community, or for user services. These functions were assigned to an archive data center, thus freeing the Mission Data System from responsibility for these essential functions that are best conducted by professional data managers.

The SIS study and the data center survey provide some guidelines for planning NewDISS costs. Based on the survey, there is some expectation that the NewDISS model of distributed, heterogeneous and cooperative data centers may be more cost-efficient than the one-size-fits-all approach of ECS. Regardless of this putative cost-efficiency, NASA's ESE must maintain adequate funding for data system and mission ground system development and operations. Using a conservative guideline and based on the SIS study (Table E-1), future NASA ESE missions should expect to spend around 10–25 percent of their resources on the data system's development, operations, and services and roughly another 10 percent on their mission operations system. Mission Teams that propose expenditures of funds less than these percentages must be selected with extreme caution.

While it is possible at this point to provide a rough guideline for the funding required for the new science data centers, it is more problematic to calculate the necessary funding for the Backbone Data Centers, and Mission and MultiMission Data Systems. To do so adequately, the kinds of missions to be flown, and therefore the volume, number of products, and types of processing needed, must be considered. Estimates will also need to be made of the funds required for developing new data centers or to retrofit existing data centers to meet the new requirements.

As a next step, the authors of this report recommend that NASA, perhaps using the NewDISS transition team recommended in section E.6, perform a more thorough cost analysis of the NewDISS. The success of the analysis will be predicated on first expanding and refining the NewDISS model. Secondly, the implementation team will need to elaborate further the functions and costs of NewDISS components, including the data centers and infrastructure components necessary for a complete NewDISS.

E.5 Metrics for Success

Other parts of this document have addressed the issues of the components, standards and practices needed to enable open participation in NewDISS. NewDISS management must ensure that it has success criteria to measure how data are used and how capabilities are being utilized. It is beyond the scope of this report to completely define this approach; however, it is assumed that NASA will manage the NewDISS investigators, institutes and organizations by defining management practices in four general categories: financial/accounting practices, interface requirements, overall project requirements, and requirements for reviews/evaluations. Examples are given in appendix A of the type of NASA management directives that will help provide a standard approach and rational set of criteria for evaluating the success (or lack of it) of investigators, institutes and organizations that participate in NewDISS. It is the recommendation of the authors of this report that this list (or a similar set of criteria) be used in the evaluation of the investigators, institutes and organizations that participate in NewDISS. Furthermore, it is recommended that such reviews and evaluations be conducted with the participation of data management experts. In conducting such reviews and evaluations the failures of NewDISS can be quickly identified and ameliorated, and the successes of NewDISS can be identified and encouraged.

E.6 Transition Planning

The goal of the report was to define the future direction, framework and strategy for NASA's ESE data and information processing, near-term archiving and distribution. Clearly, it is beyond the scope of this report to determine any particular implementation or to define a specific transition from the current suite of Earth science data management activities to a NewDISS. It is the final recommendation of this report, however, that NASA must organize a transition team without delay, and that the team should be chartered with the objective of developing a transition plan, based on the findings and recommendations of this document, that would lead to the initiation of a NewDISS starting in 2003. It is further recommended that the transition team be made up of appropriate representatives from the Earth science research and applications community, the community of Earth science data managers, plus representatives from NASA and its partner agencies.

F. REFERENCES

ESE, 2000. **NASA's Earth Science Enterprise Research Strategy for 2000-2010.**

National Research Council (NRC), 2000. **Review of NASA's Earth Science Enterprise Research Strategy for 2000-2010;** National Academy of Sciences, National Research Council; National Academy Press, Washington, DC.

National Research Council (NRC), 1999. **Global Environmental Change: Research Pathways for the Next Decade,** National Academy Press, Washington, D.C.

United States Global Change Research Program (USGCRP), 1999. **Global Change Science Requirements For Long-Term Archiving, Report Of The Workshop, October 28-30, 1998, National Center for Atmospheric Research, Boulder CO.**

National Research Council (NRC), 1994. **Panel to Review EOSDIS Plans, Final Report,** National Academy Press, Washington, D.C.

National Research Council (NRC), 1994. **1993 Data Forum: A Review of an Implementation Plan for U.S. Global Change Data and Information ,** National Academy Press, Washington, D.C.

Committee on Earth and Environmental Sciences, 1992. **U.S. Global Change Data and Information Management Program Plan,** National Science and Technology Council, Washington, D.C.

(An NRC review of the initial plans for GCDIS is described in **The 1992 Data Forum: A Review of a Federal Plan for Managing Global Change Data and Information,** National Academy Press, Washington, D.C., 1992.)

National Research Council (NRC), 1992 **The 1992 Data Forum: A Review of a Federal Plan for Managing Global Change Data and Information,** National Academy Press, Washington, D.C.

National Research Council (NRC), 1991. **Solving the Global Change Puzzle: A US Strategy for Managing Information,** Committee on Geophysical and Environmental Data, National Academy Press, Washington, D.C.

Committee on Data Management and Computation (CODMAC), National Research Council (NRC), 1988. **Selected Issues in Space Science Data Management and Computation,** National Academy Press, Washington DC.

Committee on Data Management and Computation (CODMAC), National Research Council, 1986. **Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences,** National Academy Press, Washington, D.C.

APPENDIX A

Examples of NewDISS Management Practices and Requirements

Other parts of this document have addressed the issues of the standards and practices needed to enable open participation in NewDISS. This appendix goes into somewhat more detail on the management practices that need to be considered in minimizing the risks associated with development and delivery of NewDISS components.

Overall, it is assumed that NASA will manage the NewDISS investigators, institutes and organizations by defining management practices in four general categories: financial/accounting practices, interface requirements, overall project requirements, and requirements for reviews/evaluations.

Examples are given below of the type of NASA management directives that will help provide a standard approach and rational set of criteria for evaluating the success (or lack of it) of investigators, institutes and organizations that participate in NewDISS. Furthermore, it is recommended that issues listed below be specifically addressed in any Announcement of Opportunity that is released by NASA for NewDISS.

Financial/Accounting Practices

- Description of the funding profile.

- Description of the required technical, managerial and financial reports.

- How often does NASA want the reports and their format?

- Agreement on whether the PI will be able to hold reserves or will the reserves be held by NASA?

- If NASA holds the reserves, a description of the process for transferring reserve funds to the PI.

Interface Requirements

- Description of the responsibility (if any) the PI has in supporting the integration of the software that is delivered to NASA.

- Description of the Level 0 and Level 1 interfaces including points of contact (organizations).

- A list of all the available Level 0 ancillary files (needed for processing the data).

- Procedures describing how dependencies will be dealt with:

- Data from other PI's.

- Data from other instrument PI's.

- Ancillary data.

- Data from other agencies.

- Transfer of the archived data to the DAAC/long-term archive.

- Description of the catalog interfaces through which data can be identified

- Description of distribution requirement.

- Distribution schedule (how often does the data need to be distributed?)

- How many users are to be supported?
- Network access mechanisms
- Distribution media.
- Charging for production and distribution costs.

Requirements for Reviews/Evaluations

Description of the required/mandatory reviews, their format and tentative location.

Milestones tracked by NASA for the following activities:

- Software development.
- Processing system development.
- Archive and distribution systems development.
- Operations.

Overall Project Requirements

If NASA wants a copy of the processing software

- When does it need to be delivered?

- How many versions need to be delivered?

- What documentation needs to accompany the software?

A list of data products that NASA wants delivered to an archive facility (levels of data).

- A schedule for the deliveries.

- Format of the data.

A list of all the required documentation and which documents are deliverables to NASA.

A list and description of the mandatory metadata.

A list of the required data formats and their description.

Description of the coding standards that a PI must follow.

Description of the mandatory practices.

Description of the requirements for platform independence.

Description of any standard hardware or software.

Description of the mandatory science QA requirements.

Description of PI processing responsibilities.

- During the mission.

- After the mission ends.

- Required processing schedule.

Description of archiving responsibilities.

- During the mission.

- After the mission ends.

Description of user support.

Description of required display/visualization tools.

Testing and acceptance requirements.

APPENDIX B

Space Science Data Management

The goal of the Office of Space Science (OSS) data management activity is a coherent and coordinated data environment providing scientists, educators, and the general public timely, expedient access to high-quality space science data holdings acquired through flight missions and investigations. The basic operating principles for achieving that goal, as well as the trends, challenges, and plans for evolving the data environment are discussed below, which then concludes with some summarizing “lessons learned” in the process to date.

Basic Operating Principles

- The space science archive data center infrastructure is organized by science discipline or topic (e.g., astrophysics Science Archive Research Centers, Planetary Data System, etc.), with active involvement of the science user in the management and operation of the data services. Establishment of new nodes, as well as continuation of existing nodes is determined via science-driven peer review.
- The Space Science Data System Technical Working Group (SSDS TWG) is responsible for coordinating data services across the archive data center infrastructure and for enhancing the integration, interoperability, and commonality of such services.
- Each OSS mission will plan for the total activity associated with the flow of science data, from acquisition, through processing, data product generation and validation, to archiving and dissemination. This is documented in a Project Data Management Plan (PDMP).
- OSS missions will coordinate data management planning with the appropriate space science archive data center for interface requirements, data formats, and other standards for delivery of data products, and those arrangements are reflected in the PDMP.
- The OSS Senior Review process consists of periodic review of operating space science missions, systematically applied across all space science disciplines, for the purpose of determining priorities for responding to requests for budget augmentations and/or extension of operating lifetime. Included as one of the evaluation factors is the timely delivery of data to archives, as well as accessibility and usability of data and related software. The Senior Review process is also used to review, evaluate, and determine priorities for the archive data center infrastructure.

Trends and Challenges

- Multitude and diversity of missions

The number of simultaneously operating space science missions will continue to grow into the future (see figure 1), with a corresponding growth in sheer volume of science data flowing into archives. The mission type is also quite diverse, ranging from large observatory class missions (e.g., Chandra X-Ray Observatory, SIRTf) to multiple Small Explorer- and Discovery-class missions to PI-driven missions operated from their institutions.

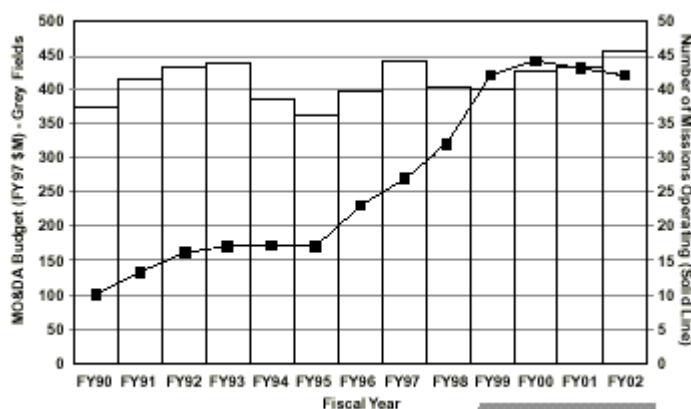


Figure 1

- Continued budget pressures

The OSS Mission Operations and Data Analysis (MO&DA) budget line funds mission operations, science operations (including data processing, data distribution, and data archiving), and science data analysis. Total funding support for operating missions will remain roughly constant into the future. The combined effect of this trend with the above results in a dramatic decrease in average MO&DA funding per mission, which will continue to decline into the future (see figure 2). The total OSS funding for archive data center infrastructure will also be held roughly constant into the future at \$35M per year. So the pressure will continue to grow, evolve and operate the data archiving infra-structure at minimum costs.

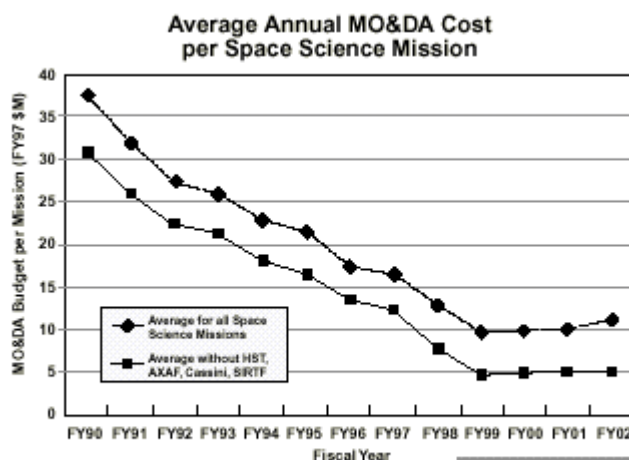


Figure 2

- Increasingly interdisciplinary nature of the space science research

From understanding the mechanisms of solar variability and the specific processes by which the Earth and other planets respond, to the concept of a digital sky for simultaneously

“observing” the sky in all wavelengths, the nature of the research collaboration often spans multiple space science disciplines. The challenge for the data archive infrastructure is to be perceived as a comprehensive, collective whole, allowing transparent user view paths as driven by the particular multidiscipline collaboration for heterogeneous searches, queries, and fusion requests.

- Robust Data Infrastructure

There are many challenges in evolving a robust infrastructure of widely distributed data resources. There are all the standard and interface issues for maintaining and enhancing interoperability, commonality, sharing, etc. Extensibility and scalability are factors as requirements and capabilities age and evolve. And then there is the whole set of issues associated with exploiting such an explosive information technology area and infusing new technology across the federated data union.

Evolution of the Space Science Data System

As noted earlier, the current space science data infrastructure is primarily organized by science discipline and consists of the elements listed below.

- The National Space Science Data Center (Goddard Space Flight Center, Greenbelt, MD)—Permanent archive for space science data, and principal archive site for space physics data and astronomical catalog data
- Solar Data Analysis Center (Goddard Space Flight Center, Greenbelt, MD)
- Planetary Data System, with the following components:
 - Central Node (Jet Propulsion Laboratory, Pasadena)
 - Atmospheres (New Mexico State University, Las Cruces)
 - Rings (Ames Research Center, Mountain View, CA)
 - Small Bodies (U. of Maryland, College Park)
 - Planetary Plasmas (U. of California at Los Angeles)
 - Geosciences (Washington University, St. Louis, MO)
 - Imaging (U.S. Geological Survey, Flagstaff)
 - Astrophysics Science Archive Research Centers
 - High-Energy Astrophysics (Goddard Space Flight Center, Greenbelt, MD)
 - Infrared Astronomy (Infrared Processing and Analysis Center, Pasadena)
 - Ultraviolet/Optical Astronomy (Space Telescope Science Institute, Baltimore)

A user group has recommended an architecture for a formal Sun-Earth Connection Data System. This will complete the coverage of archive data infrastructure across all the space science research disciplines. Implementation has been initiated and components will be competitively selected.

Interoperability has been enhanced over the past year by providing browsing and location capability across the science data holdings from any and all of the disparate user interfaces for the archive data centers. This builds on a standardizing front-end modifier, “AstroBrowse,” developed within the astrophysics data centers, which is now being extended

to the other discipline components. The next step will then be to actually correlate requests for the multiple datasets located across the multiple data sites, merge or fuse the results and deliver the data to the user, through a consistent and familiar user interface.

It is important to note that the Space Science Data System (<http://ssds.nasa.gov>) will evolve under the auspices of the SSDS TWG. This group consists of representatives of all the space science data providers and users, as well as data system technologists.

Summarizing Lessons Learned

The active involvement and ownership by the science community are absolutely crucial conditions for success.

Build and evolve incrementally—AVOID LARGE DEVELOPMENTS.

Minimize any separation between “developers” and science data providers and users.

APPENDIX C

List of Acronyms

AO—Announcement of Opportunity
DAAC—Distributed Active Archive Center
ECS—EOSDIS Core System
EOSDIS—Earth Observing System Data Information System
ESE—Earth Science Enterprise
ESIP—Earth Science Information Partnership
GOS—Global Observing Strategy
IDS—Interdisciplinary Science
IGBP—International Geosphere-Biosphere Programme
IGBP-DIS—Data and Information Systems Framework Project of the IGBP
IHDP—International Human Dimensions Program
IMS—Information and Management System
LIS—Lightning Imaging Sensor
MO&DA—Mission Operations and Data Analysis
NewDISS—New Data and Information Systems and Services
NGI—Next Generation Internet
NOAA—National Oceanic and Atmospheric Administration
NSIDC—National Snow and Ice Data Center DAAC
OES—Office of Earth Science
OSS—Office of Space Science
PDMP—Project Data Management Plan
SCF—Science Computing Facility
SeaWiFS—Sea-viewing Wide Field-of-view Sensor
SEDAC—Socioeconomic Data and Applications Center
SIRTF—Space Infrared Telescope Facility
SSDS TWG—Space Science Data System Technical Working Group
TSDIS—TRMM Science Data and Information System
USGCRP—U.S. Global Change Research Program
WCRP—World Climate Research Program
WP-ESIP—Working Prototype Earth Science Information Partners